CS TRACK
Investigating Citizen Science

Horizon 2020 / Science with and for Society Programme
Grant agreement number: 872522

Models to identify background factors associated with the CS activity

D4.3

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872522

Title of project	CS Track
Full title of project	Expanding our knowledge on Citizen Science through analytics and analysis
Title of this document	Models to identify background factors associated with the CS activity
Number of this document	D4.3
Dissemination level	Public
Due date	30.11.2021
Actual delivery	12.1.2022
Versioning history	
Authors	Ohto Sabel, Aaron Peltoniemi, Heli Kauppinen, Kari Nissinen, Joni Lämsä, Emilia Lampi, Raija Hämäläinen (Figure graphics of Figures 1,2,4,5, and 6 by URJC -team)
Executive summary	This deliverable is based on the analysis of the CS Track survey. It summarizes the characteristics of citizen scientists, discusses how learning and knowledge building occurs and explicates the structure and basis needed to build models that illustrate the associations between background factors and different forms of citizen science activities. The models provide background information regarding the future of CS Track, and, together with other CS Track data, are to be utilized in determination of accreditation practices for CS and thus the forming of policy recommendations.

Table of Contents

1. Introduction

CS Track aims to understand citizen science (CS) in a comprehensive way by inviting citizen scientists engaged in various CS projects to deepen and broaden our knowledge of CS activities and their impacts. Acquiring such knowledge will allow individual citizens, organisations and societies to determine effectively the potential benefits and caveats of CS. In WP4, we focus on a research-based evaluation of the impacts of CS activities with higher-level, multi-perspective analysis (see D4.1 and 4.2 for detailed descriptions). D4.3 is grounded on the work reported in D4.1 and D4.2 as well as based on the close collaboration with other work packages (especially WP1, 2 and 3). Next, we provide a general overview of how the CS Track project creates a novel understanding on CS activities, based on the new set of analytical methods. At the end of the document, we illuminate the first-hand version of the models used to identify factors explaining the variation in CS activities based on theoretical assumptions and empirical support (the final version of the models will be published in a peer-reviewed article based on the CS Track publication plan). Overall, the specific aim of D4.3 is to introduce the way we build models for identifying how background factors are associated with CS activities as well as to illustrate first-hand results of these models.

We use the data from the CS Track survey (N=1,083) to provide indicators regarding the backgrounds of citizen scientists, the different forms of participation as well as the learning and knowledge building processes in CS to shed light on the range of factors that contribute to sustainable CS activities and thus successful projects. The survey itself was based on theoretical and empirical research with the goals of broadening our understanding of CS, gaining general information about CS projects, participation of citizen scientists and their ways of engagement in CS, learning, knowledge building and scientific learning in CS, and the diversity of citizen scientists (see D 4.2; Paajanen et al., 2021).

Since CS is a highly multilayered concept, we selected theoretical background indicators for our survey in close collaboration with CS Track partners (see D4.1 and D4.2). Data was collected from 19.1.2021 until 12.8.2021 through the web-based survey environment of Webropol, and the data was refined and organised by two statistician experts. Then, we applied multiple statistical methods (e.g. non-parametric tests, and correlation and regression analyses) to investigate the potentials and limitations of the data. In the following paragraphs, we introduce the general overview of the analysis behind the models.

In practice, building the models needed for explaining the variation in CS activities were implemented in two stages. *The first phase of model building is based on the state-of-the-art D1.1 (and D1.2) by WP1 as well as the socially-situated nature of CS* (i.e. the CS activities are always related to the personal lives and history and real-life problems of citizen scientists). We referred to the previous research on CS as a starting point for analysing the survey data (see Ceccaroni et al., 2017). In short, previous studies on CS provided support in studying the common factors for participation in CS projects. In practise, the following four characteristics were studied:

First, socio-demographic background revealed the diversity of citizen scientists.

Second, the activities, engagement, motivation and forms of participation of citizen scientists were studied by delving into the various forms and ways of interaction, collaboration, sharing and engaging in project activities.

Third, learning and knowledge building were examined by viewing the activities for experienced learning and knowledge building practises of citizen scientists as well as their understanding of science and the scientific process.

Fourth, the reasons for participation in CS were studied by reviewing the following factors: interest in a theme or topic, desire to help, social reasons, status rewards, personal growth, involvement in a community, contribution to scientific research, values, enjoyment, opportunities to learn and to share knowledge, and possible career benefits (see also, D4.1 Lampi et al., 2020).

Further descriptions are presented in the manuscript regarding "how citizen scientists learn" submitted for peer review as shortly described below. The description of essential perspectives can be found in three eMagazine articles that have been published this year:

- Typical characteristics of citizen science (CS) participants. (Peltoniemi et al., 2021a) https://cstrack.eu/format/graphical-article/typical-characteristics-of-citizen-science-cs-participants/

- What technical devices/platforms are used most by Citizen Scientists in their projects? (Peltoniemi et al., 2021b) https://cstrack.eu/format/graphical-article/what-technical-devices-platforms-are-used-most-by-citizen-scientists-in-their-projects/

- Knowledge building and roles in Citizen Science – findings from the CS Track survey 2021. (Peltoniemi et al., 2021c) https://cstrack.eu/format/graphical-article/knowledge-building-and-roles-in-citizen-science-findings-from-the-cs-track-survey-2021/

The first layer of analyses as presented in the eMagazine articles used descriptive statistical methods to study the characteristics of citizen scientists (see, section 2; Peltoniemi et al., 2021a). Regarding how citizen scientists learn and build knowledge, we examined Pearson correlation coefficients between the activities to create a learning situation entity. Through this correlation comparison, we identified, in total, three different learning situation entities. Next, we examined how the project-related and personal factors were associated with the perceived learning in these three different entities (a Kruskal-Wallis test and Dunn's A test were used). (A detailed description of methods and results are presented in the manuscript that was submitted for peer review on 17th of December. However, please note that since this D4.3 is a public document, only a general level description is presented to ensure a blind-review.)

The second phase of model building focused on creating a deeper understanding of the association between the socially-situated nature of CS with selected CS activities. In contrast to the first phase, statistically significant background factors (see Figure 3) were investigated in a more detailed level, reflecting a more bottom-up and data-driven approach. This allowed us to analyse and observe the current state of critical activities with empirical evidence. In practice, our aim was to find new ways to identify successful links related to CS activities. The statistical method used in the analyses is a logistic regression (see e.g. (Hämäläinen et al., 2015, 2019, 2021)). In the next sections, figures 3 and 7 provide a general overview of the variables entered in each model.

2. Who are citizen scientists?

To develop further understanding of how the background factors (see Figure 3) are associated with different forms of CS activities, first, we focused on the current demographics of citizen scientists through examining the data from the CS Track survey (N=1,083). Regarding some typical characteristics of citizen scientists, we found that citizen scientists "are more likely to be male (see Figure 1) and married but living without children in a small city or large town (Peltoniemi et al., 2021a)." In addition, it was not uncommon for citizen scientists to possess a scientific background such as a master's degree (Peltoniemi et al., 2021a). In the following table (Table 1), we present the main results and publication details for each background factor. At the end of this section, Figure 3 is provided to illustrate the statistically significant background factors that form the grounding elements for future analysis.

Background factor	**Main result (and thus grounding future work)**	**Publishing details**
Age	Four percent were between the ages of 16 and 20 years old, 43% were between the ages of 21 and 50 years old, and 53% were over 51 years old.	(Published eMagazine: Peltoniemi et al., 2021a, 2021d)
Gender	Fifty-six percent were male, 41% were female, and the remaining 4% of were non-binary or preferred to self-describe or not to answer the question.	(Published eMagazine: Peltoniemi et al., 2021a, 2021d)
Education	Fourteen percent had a bachelor's level, 36% had a master's level, and 17% had a	(Published eMagazine:

	doctoral level of education. Only 3% had a lower secondary level or only a primary level of education.	Peltoniemi et al., 2021a, 2021d)
Employment	Fifty percent were employed, 26% were pensioners, and 9% students.	(Forthcoming eMagazine)
Income	Twenty-one percent had a yearly income greater than €70,000, and 54% had income between €10,000 and €69,999. Only 7% had an income less than €10,000. Nearly one-fifth (19%) did not report their income level in the survey.	(Peltoniemi et al., 2021d; Hästbacka et al., 2021)
Household	Forty-five percent were living in a two-adult household and 19% in a single household.	(Forthcoming eMagazine + Hästbacka et al., 2021)
Children	Seventy-two percent of households did not have children, 11.5% had one child and 11.8% had two children.	(Forthcoming eMagazine + Hästbacka et al., 2021)
Area of residence	Most lived in a small city or large town (26%), a medium city (21%), or in a large city (16%).	(Forthcoming eMagazine + Hästbacka et al., 2021)
Marital status	Fifty-five percent were married, 13% were single and 10% responded with “other.”	(Forthcoming eMagazine + Hästbacka et al., 2021)

Table 1. Main results of background factors and their publication details.

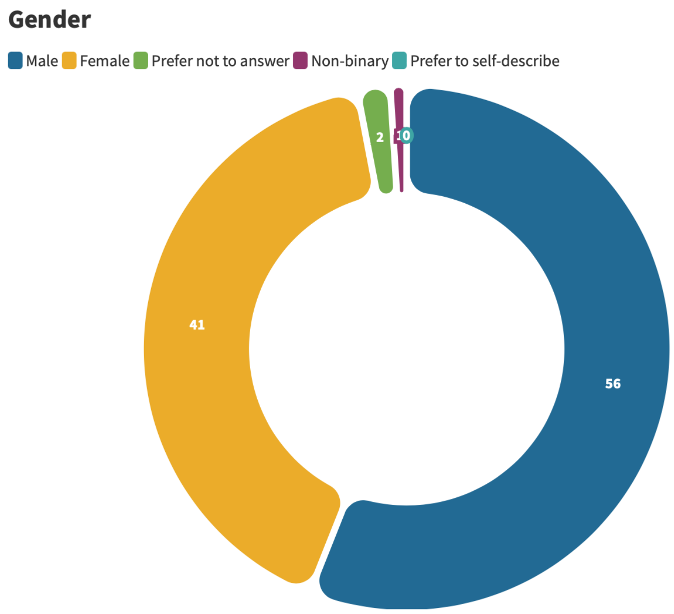

Figure 1. Gender distribution of citizen scientists (N=1,083)

For the future, our analysis opens several novel perspectives on understanding CS. For example, while debated, recent demographic research in CS highlights that citizen scientists tend to be middle-aged men (see Figure 1 above) who are interested in science and possess a university level of education (Cooper et al. 2021; Pateman, Dyke & West 2021; Vasiliades et al. 2021). In our sample, 36% of citizen scientists had an educational background equivalent to a master's degree. In addition, only 3% of citizen scientists had a lower secondary education or lower (see Figure 2 below).

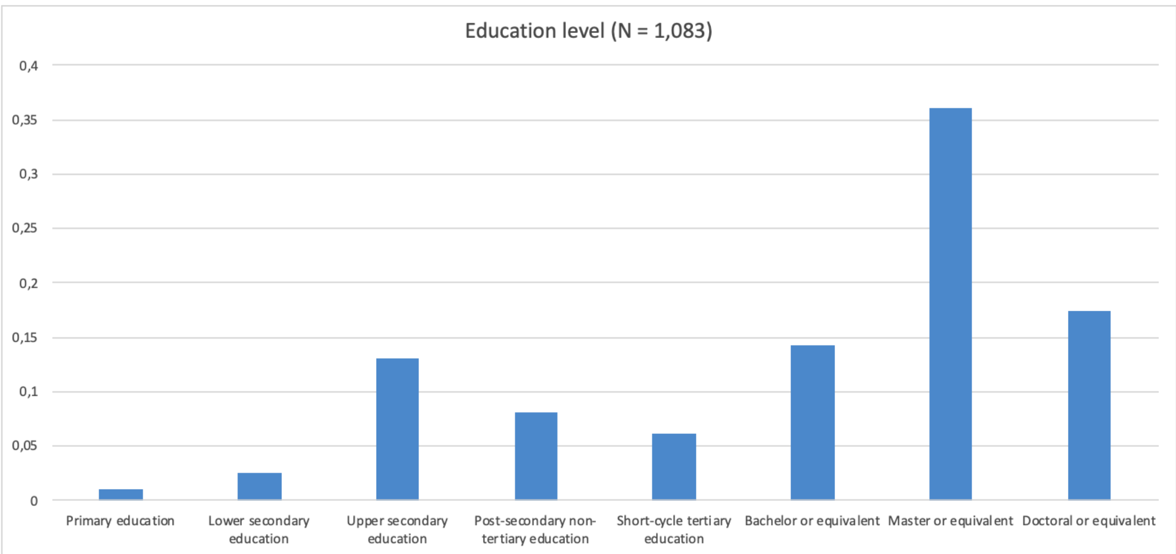

Figure 2. Education level of citizen scientists

However, as later illustrated in section four, our models provide a more insightful understanding of this phenomenon. In particular, it seems that the roles adopted by citizen scientists in CS projects were more significant and thus relevant indicators (than gender and educational level) regarding learning and participation duration in CS. The participation and retention of citizen scientists are also important for the sustainability of CS projects (Roche et al., 2020; Vohland et al., 2021). In our sample, we noticed that 41% of citizen scientists have been active in CS for more than 10 years. Citizen scientists who indicated engagement in CS from zero to one years was 17%. These findings are considered in our models for further analysis (see Figure 3 below).

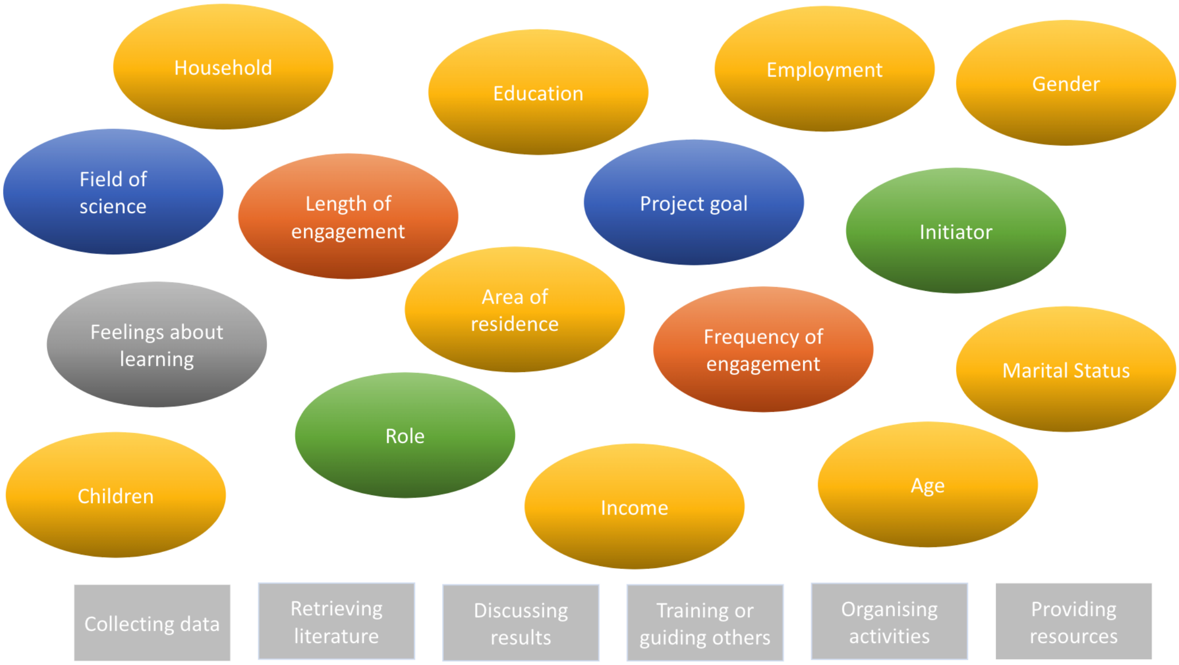

Figure 3. Background factors (ellipses) and CS activities (rectangles) for future statistical modelling.

Based on our statistical analysis, a set of background factors (multi-coloured ellipses in Figure 3) were selected for statistical modelling. The background factors include: (a) socio-demographic factors (yellow ellipses) age, gender, education, employment, income, household, children, area of residence and marital status, (b) length and frequency of engagement (red ellipses), (c) the field of science and project goal (blue ellipses), (d) the role in CS projects or the initiator of CS activities (green ellipses), and (e) feelings about learning in CS (grey ellipse). These background factors are used for modelling six different CS activities (grey rectangles in Figure 3). The six modelled activities are: 1) collecting data, 2) retrieving (scientific) literature, 3) discussing results, 4) training or guiding others, 5) organising activities, and 6) providing resources (see section 4 for further details).

3. How do learning and knowledge building occur?

The objectives found in CS projects usually address pertinent as well as local topics of interests such as water quality in collaboration with citizen scientists (Roche et al., 2020, Vasiliades et al., 2021; Vohland et al., 2021). Learning and knowledge building processes found in CS projects are quite diverse and include formal learning (e.g. scientific literacy) as well as informal learning (e.g. interpersonal skills) opportunities (Aristeidou & Herodotou, 2020; Roche et al., 2020).

Our survey of citizen scientists identified 12 activities in CS projects that enable learning. The third eMagazine article provides a general overview of how citizen scientists participated in different kinds of knowledge building activities (see Peltoniemi et al., 2021c). Figure 4 below highlights the activities in which learning was most often reported. Specifically, we found that citizen scientists were able to learn the most by interacting with others or information online as well as by reflecting on prior knowledge. Our results underscore the research (e.g. Phillips et al., 2018) that the learning goals of CS projects are not necessarily congruent with the experiences of citizen scientists.

Figure 4. Experienced learning and mean scores for activities in CS projects (N=1,083)

We also investigated if and how the roles that were adopted (e.g. citizen scientist, facilitator, project coordinator, etc.) are associated with experienced learning and knowledge building (see, Figure 5; Peltoniemi et al., 2021c). From Figure 5, we can see that citizen scientists felt like they had learned the most while searching for information from the Internet and the least when creating something new (column 1, rows 10 and 2 respectively). Those who adopted leadership roles, such as being the leader of a CS project, felt like they had learned the most while talking and interacting with others but the least when reading professional magazines and journals (column 5, rows 12 and 7 respectively). Finally, those who adopted roles that were not specified on the survey seemed to experience the most learning when engaging in activities that were not listed in the survey (column 7, row 1).

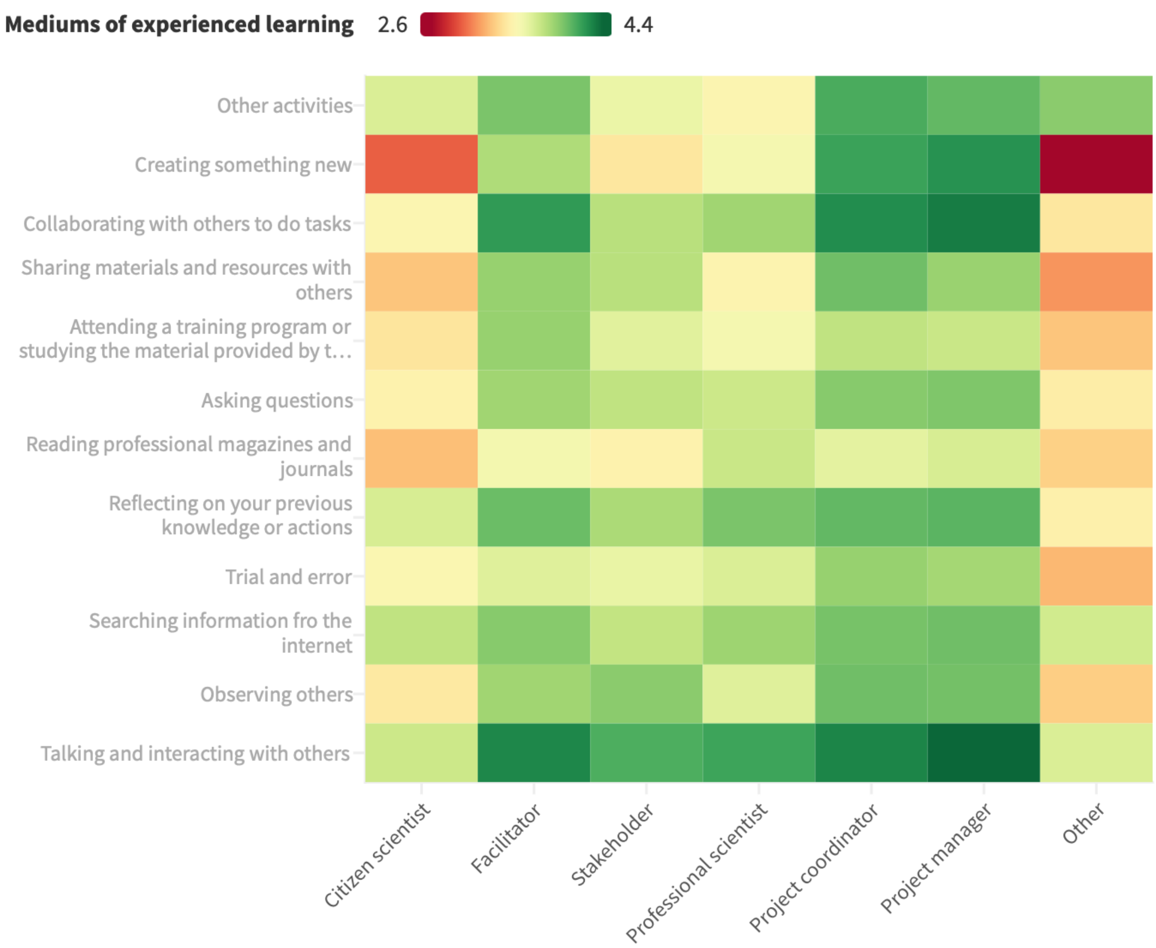

Figure 5. Experienced learning for adopted roles across different activities (N=1,083)

There is also an ever-evolving technological landscape that challenges and changes the learning and knowledge building experiences of citizen scientists. For example, citizen scientists may need to stay updated in social media channels and develop their own digital competences, or to adapt their activities with the potentials and challenges of the digitalised world (e.g. novel forms of data collection). Therefore, in addition to learning and knowledge building, we briefly investigated how technologies were used in CS. Thus, our second eMagazine article outlined how technical tools were broadly used in CS projects (see Peltoniemi et al., 2021b). Of the 12 different tools and platforms that are used in CS activities, the top three were websites, smartphones and databases. Citizen scientists indicated that they used technology primarily for gathering scientific data and knowledge as well as for communication and dissemination of information with others and researchers (Figure 6 below and Peltoniemi et al., 2021b). This analysis shall be continued together with RIAS, UPF and WLW towards creating a peer-reviewed publication based on the publication plan of the project (agreed and initiated in July 2021).

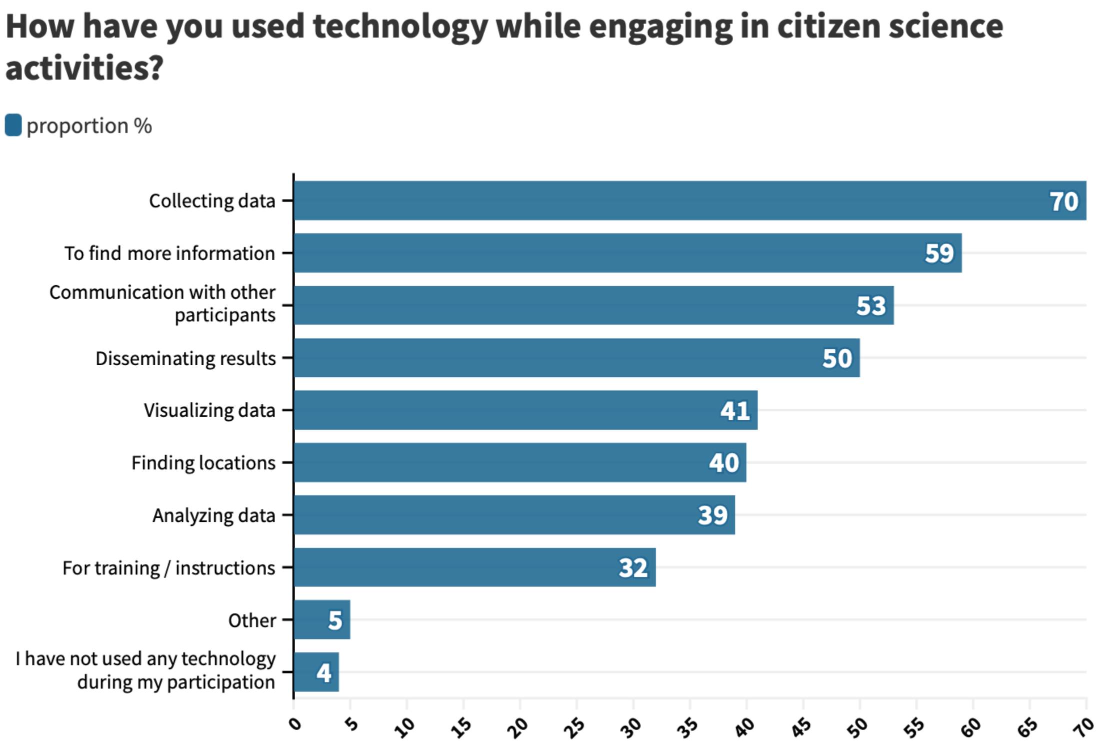

Figure 6. The use of technology while engaging in CS activities (N=1,083)

4. Building models to illustrate the association between background factors and different forms of CS activities

In this section, we introduce the models in which the variation in CS activities are explained based on the background factors (see section 3). The models are based on theoretical assumptions (e.g. D1.1) and empirical support. From the survey data (N=1,083), we selected six different CS activities that illustrate the diversity of CS projects in terms of whether citizen scientists had reported their participation in the 1) collection of data, 2) retrieval of scientific literature, 3) discussion of results, 4) training or guiding of others, 5) organising activities, or 6) providing resources such as computing power of their computers to solve complex simulation problems.

For each activity, we conducted a binary logistic regression analysis to examine which project-related and personal factors explain (and predict) the probability of participating in it. The statistical significances of background factors were determined by the Chi-squared test, and the statistically significant (p value less than 0.05) factors were only included in the models. The explanatory power of each considered model was examined with appropriate R-squared statistics.

Altogether, we focused on the following 11 different factors: (1) How long citizen scientists have been engaged in CS, (2) how often do they engage in CS, (3) how do they feel about learning in CS, (4) who was the initiator of the CS project, (5) what was the field of science of the CS project, (6) what is the goal of the CS project, (7) the role (in the CS project), (8) age, (9) gender, (10) education, and (11) income of citizen scientists.

A detailed description of methods and results are presented in the manuscript that will be submitted for peer review in Spring 2022 based on the publication plan. However, please note that since D4.3 is a public document, only a general illustration of the method and models are presented below to ensure a blind-review. In the forthcoming article, we present a model for each of the six activities (in total, six models) so that the statistically significant predictors are included and presented (see Figure 7).

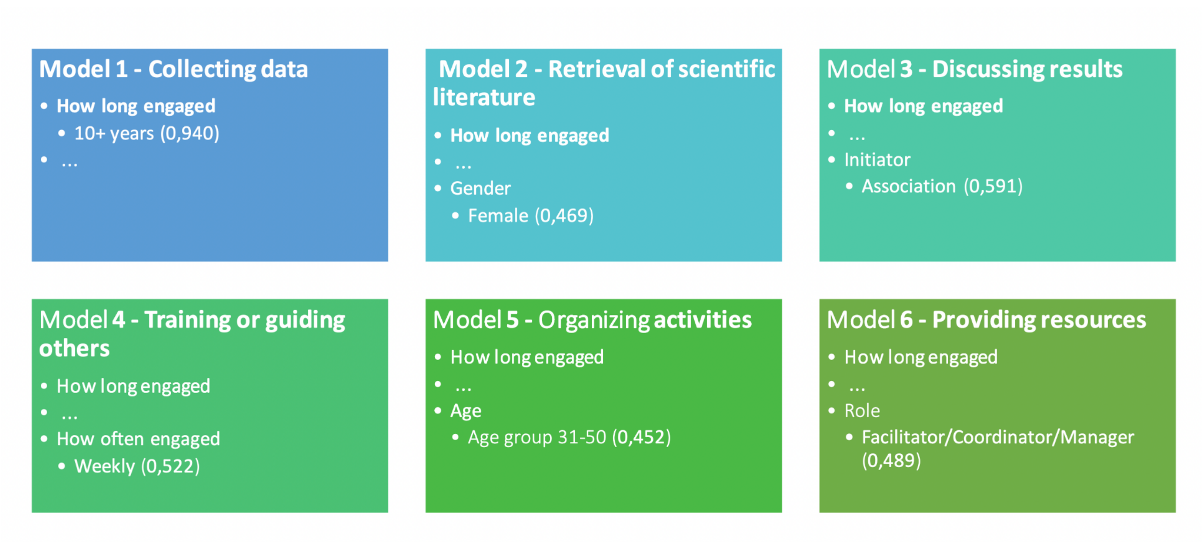

Figure 7. An illustration of statistical modelling with example significant explanatory variables

In Figure 7, we illustrate how our statistical modelling will construct an explanatory model of a particular phenomena. For example, the factor "How long the respondent has been engaged in CS activities", was significant in all six models. That is, it was significantly related with all six activities in question. Further, if the respondent had been engaged in CS for 10 years or more, we can estimate that he/she had participated in data collection activities with probability of 0.940, according to Model 1. Figure 7 shows only a fragment of results, but it highlights some of our final findings (which are to be published in a peer reviewed journal). One such finding was that "how long engaged" was systematically associated with all six activities, while factors like "age" and "gender" did not appear significant in all six models.

5. Discussion

To conclude, the data from the CS Track survey (N=1,083) was analyzed to develop further understanding of how different factors (e.g. gender, age, roles in CS) are associated with different forms of participation in CS activities. Our aim was to develop a novel understanding of citizen scientists' subjective experiences of learning, knowledge building activities and processes in CS projects. We explored the background factors related to experienced learning and participation of citizen scientists and their activities. Our models are based on the theoretical background of D1.1 and the socially-situated nature of CS. Thus, the ultimate goal of these models is to understand and link these factors together to create meaningful and relevant interpretations. In practice, we identified factors explaining the variations of citizen scientists in CS activities based on theoretical assumptions as well as empirical support. Next, we briefly elaborate some of the interesting findings that will provide the ground for future work of CS Track.

First, researchers have noted that CS shares similarities with volunteerism, particularly, since citizen scientists are typically those who have the flexibility to perform unpaid work (Vasiliades et al., 2021; Vohland et al., 2021). When examining the sample of our survey, we noticed that citizen scientists, even if married, usually live in a household of two in a small city or large town. When citizen scientists indicated their income level, 30% had an income between €10,000 – €39,999 and only 21% had an income from €70,000 or above (Peltoniemi et al., 2021d).

Second, the first-hand results from the survey data indicated that citizen scientists have possibilities to engage in a variety of CS activities, from organizing projects and training others to collecting data and discussing achieved results. Even though the different types of projects enable different forms of participation (e.g. depending on the initiator, goals, or research area of the project), the personal experience of citizen scientists seems to play an important role. Namely, we found that the duration and frequency of participation in CS projects were significant predictors concerning the participation of citizen scientists in different CS activities.

Third, the observed diverse forms of participation may also be associated with learning (e.g. retrieval of scientific literature as a CS activity) and knowledge building (e.g. discussing results as a CS activity) that take place in CS projects. Many citizen scientists also utilise various technological tools that provide possibilities and forms of participation that were previously more difficult to conduct (Aristeidou & Herodotou, 2020; Vasiliades et al., 2021). For example, digital databases provide access to scientific literature and various tools thereby easing the collaboration among citizen scientists, and thus enhancing discussions on the results of the project and the guidance of others.

In summary, this report provides background information for the future of CS Track. Furthermore, our models identify factors explaining the variations of citizen scientists in CS activities. This knowledge shall be investigated together with other CS Track data (see, WPs 1,2, and 3) and applied to determine accreditation practises for CS (D4.5) and in forming the policy recommendations (D4.4). Finally, the characteristics of citizen scientists in CS projects, their specific orientations towards learning, knowledge building, science and scientific process will support methodological triangulation (in collaboration with CS Track partners) of other empirical data to enhance and deepen our understanding of the world of CS. For example, as described in D1.2 the workshop held November 11/12 in Jyväskylä was dedicated to intensifying and advancing the triangulation work in CS Track. Collaboration (between JYU, UPF and RIAS) is foreseen towards integrating the database (WP2), the survey data (WP4) and the data-intensive computational analyses (WP3).

References

Aristeidou, M., & Herodotou, C. (2020). Online citizen science: A systematic review of effects on learning and scientific literacy. Citizen Science: Theory and Practice, 5(1): 11, 1–12. DOI: https://doi.org/10.5334/cstp.224

Ceccaroni, L., Bowser, A., & Brenton, P. (2017). *Civic Education and Citizen Science: Definitions, Categories, Knowledge Representation* [Chapter]. Analyzing the Role of Citizen Science in Modern Research; IGI Global. https://doi.org/10.4018/978-1-5225-0962-2.ch001

Cooper, C. B., Hawn, C. L., Larson, L. R., Parrish, J. K., Bowser, G., Cavalier, D., Dunn, R. R., Haklay, M., Gupta, K. K., Jelks, N. T. O., Johnson, V. A., Katti, M., Leggett, Z., Wilson, O. R., & Wilson, S. (2021). Inclusion in citizen science: The conundrum of rebranding. Science, 372(6549), 1386–1388. DOI: https://www.sciencemag.org/lookup/doi/10.1126/science.abi6487

Hämäläinen, R., De Wever, B., Malin, A., & Cincinnato, S. (2015). Education and working life: VET adults' problem-solving skills in technology-rich environments. Computers & Education, 88, 38–47. https://doi.org/10.1016/j.compedu.2015.04.013

Hämäläinen, R., De Wever, B., Nissinen, K., & Cincinnato, S. (2019). What makes the difference – PIAAC as a resource for understanding the problem-solving skills of Europe's higher-education adults. Computers & Education, 129, 27–36. https://doi.org/10.1016/j.compedu.2018.10.013

Hämäläinen, R., Nissinen, K., Mannonen, J., Lämsä, J., Leino, K., & Taajamo, M. (2021). Understanding teaching professionals' digital competence: What do PIAAC and TALIS reveal about technology-related skills, attitudes, and knowledge? Computers in Human Behavior, 117, 106672. https://doi.org/10.1016/j.chb.2020.106672

Hästbacka, H., Lampi, E., Sabel, O., Lämsä, J., Oksanen, U., & Hämäläinen, R. (2021). *Preliminary Survey Findings for D4.2*.

Lampi, E., Lämsä, J., & Hämäläinen, R. (2020). *D4.1: Towards Understanding the Forms of Participation and Knowledge-Building Activities in Citizen Science Projects*.

Paajanen, S., Lampi, E., Lämsä, J., & Hämäläinen, R. (2021). *D4.2: White* *paper: Themes, objectives and participants of CS activities*.

Pateman, R., Dyke, A., & West, S. (2021). The diversity of participants in environmental citizen science. Citizen Science: Theory and Practice, 6(1): 9, 1–16. DOI: https://doi.org/10.5334/CSTP.369

Peltoniemi, A., Hästbacka, H., Hämäläinen, R., & Lämsä, J. (2021a). Typical characteristics of citizen science (CS) participants. Citizen Science | CS Track Project. https://cstrack.eu/format/graphical-article/typical-characteristics-of-citizen-science-cs-participants/

Peltoniemi, A., Hästbacka, H., Sabel, O., Lampi, E., Lämsä, J., & Hämäläinen, R. (2021b). What technical devices/platforms are used most by Citizen Scientists in their projects? Citizen Science | CS Track Project. https://cstrack.eu/format/graphical-article/what-technical-devices-platforms-are-used-most-by-citizen-scientists-in-their-projects/

Peltoniemi, A., Kauppinen, H., Hämäläinen, R., & Lämsä, J. (2021c). Knowledge building and roles in Citizen Science – findings from the CS Track survey 2021. Citizen Science | CS Track Project. https://cstrack.eu/format/graphical-article/knowledge-building-and-roles-in-citizen-science-findings-from-the-cs-track-survey-2021/

Peltoniemi, A., Kauppinen, H., Lampi, E., Lämsä, J., Sabel, O., & Hämäläinen, R. (2021d). How Citizen Scientists Learn: Exploring Learning Perceptions Through an International Survey. *Submitted for publication*.

Phillips, T., Porticella, N., Constas, M., & Bonney, R. (2018). A framework for articulating and measuring individual learning outcomes from participation in citizen science. Citizen Science: Theory and Practice, 3(2): 3, 1–19. DOI: https://doi.org/10.5334/cstp.126

Vasiliades, M. A., Hadjichambis, A. C., Paraskeva-Hadjichambi, D., Adamou, A., & Georgiou, Y. (2021). A systematic literature review on the participation aspects of environmental and nature-based citizen science initiatives. Sustainability, 13(13), 7457. DOI: https://doi.org/10.3390/su13137457

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., & Wagenknecht, K. (2021). The science of citizen science. Cham: Springer. DOI: https://doi.org/10.1007/978-3-030-58278-4

West, S., Dyke, A., & Pateman, R. (2021). Variations in the motivations of environmental citizen scientists. Citizen Science: Theory and Practice, 6(1), 1–18. https://doi.org/10.5334/cstp.370